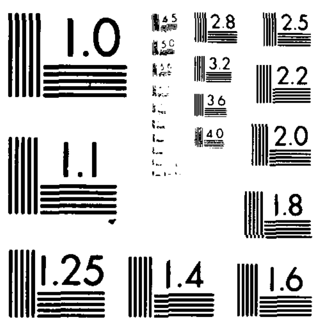
AD-A085 004 DESMATICS INC STATE COLLEGE PA F/G 12/1
OPTIMAL AUGMENTATION OF EXPERIMENTAL DESIGNS FOR ESTIMATION OF --ETC(U)
MAY 80 L A KALISH, D E SMITH N00014-79-C-0128
UNCLASSIFIED TR-112-5 NL

1 OF 1
AD A
085004

END
DATE
FILMED
6-80
DTIC



MICROCOPY RESOLUTION TEST CHART

DESMATICS, INC.

P. O. Box 618
State College, Pa. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

12

OPTIMAL AUGMENTATION OF EXPERIMENTAL DESIGNS FOR ESTIMATION OF THE LOGISTIC FUNCTION

by

Leslie A. Kalish
and
Dennis E. Smith

TECHNICAL REPORT NO. 112-5

May 1980

This study was supported by the Office of Naval Research under Contract No. N00014-79-C-0128, Task No. NR 207-037 and Contract No. N00014-75-C-1054, Task No. NR 042-334

147

TABLE OF CONTENTS

Page

I. INTRODUCTION

Two previous Desmatics technical reports [5, 9] discussed optimal designs for the estimation of the two-parameter logistic function and estimation accuracy respectively. Another report [7] discussed the use of the logistic function for prediction of impact acceleration injury. In two other reports [8, 10] injury prediction models were constructed from a set of twenty-eight $-G_x$ accelerator runs with Rhesus monkeys as subjects. It was suggested that additional runs be made in order to produce more reliable results.

This report discusses the problem of how to specify additional runs optimally. After a criterion for optimal augmentation of experimental data is presented, practical problems in its implementation are discussed. A simulation study for evaluating the augmentation procedure for the two-parameter case is described. Examples using accelerator data from [8] and [10] are given.

II. BACKGROUND

The model being used to predict injury is the logistic model, which has the form

$$P_i = f(\underline{x}_i, \underline{\beta}) = [1 + \exp(-\underline{x}_i'\underline{\beta})]^{-1} \qquad (1)$$

where $\underline{x}_i = (1, x_1, x_2, \ldots, x_k)'$ and $\underline{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)'$. P_i denotes the probability of injury on the $i^{\underline{th}}$ accelerator run, $\underline{\beta}$ is a vector of parameter values and the vector $\underline{x}_i$ gives the values of the independent (predictor) variables on the $i^{\underline{th}}$ run. The first element in $\underline{x}_i$, a "dummy variable," is included to provide for the estimation of an intercept. The $i^{\underline{th}}$ observed probability, denoted p_i is given by

$$p_i = P_i + \varepsilon_i = f(\underline{x}_i, \underline{\beta}) + \varepsilon_i$$

where ε_i denotes the error term. Because of the binomial response (injury or noninjury), p_i is either 1 or 0.

It is appropriate to use weighted least squares to estimate $\underline{\beta}$. It can be shown that the weighted least squares estimate of $\underline{\beta}$ is given by

$$\underline{b} = (\underline{X}'\underline{HX})^{-1}\underline{X}'\underline{Hy} \qquad (2)$$

where $\underline{b}$ denotes the estimated parameter vector

$$\underline{b} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)' ;$$

$\underline{X}$ denotes the design matrix (for an n run design)

$$\underline{X} = \begin{bmatrix} \underline{x}_1' \\ \underline{x}_2' \\ \cdot \\ \cdot \\ \cdot \\ \underline{x}_n' \end{bmatrix} ;$$

$\underline{H}$ denotes the diagonal weight matrix

$$\underline{H} = \begin{bmatrix} P_1Q_1 & & & & & \\ & P_2Q_2 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & P_nQ_n \end{bmatrix}$$

(where $Q_i = 1 - P_i$); and $\underline{y}$ denotes the vector of "working observations"

$$\underline{y} = (y_1, y_2, \ldots, y_n)'$$

where $y_i = (1/P_iQ_i)[p_i - P_i + P_iQ_i \ln(P_i/Q_i)]$. For more details on the estimation of $\underline{\beta}$, see Kalish and Smith [5], Walker and Duncan [11], or Smith [7].

The asymptotic covariance matrix of $\underline{b}$ is

$$\text{Var}(\underline{b}) = (\underline{X}'\underline{HX})^{-1} \qquad (3)$$

In practical applications, the covariance matrix must be estimated by

substituting

$$\underline{\hat{H}} = \begin{bmatrix} \hat{P}_1\hat{Q}_1 & & & & & \\ & \hat{P}_2\hat{Q}_2 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \hat{P}_n\hat{Q}_n \end{bmatrix}$$

for $\underline{H}$ in (3), where $\hat{P}_i = [1 + \exp(-\underline{x}_i'\underline{b})]^{-1}$ and $\hat{Q}_i = 1 - \hat{P}_i$.

For the two-parameter case (k = 1), equation (1) is a sigmoid curve of the form

$$P = \{1 + \exp[-(\beta_0 + \beta_1 x)]\}^{-1}. \tag{4}$$

LD_{100P} is defined to be the value of variable x which results in a probability P of response; that is, $P = (1 + \exp\{-[\beta_0 + \beta_1(LD_{100P})]\})^{-1}$.

III. AUGMENTATION OF EXPERIMENTAL DATA

In general, an optimal experimental design is one which "minimizes" the covariance matrix $(\underline{X}'\underline{H}\underline{X})^{-1}$. The meaning of the word minimize, when applied to a matrix, is not obvious. A number of functionals of a covariance matrix have been proposed as criteria for minimization; for each criterion there is a corresponding optimal design. (See Keifer [6].) In Kalish and Smith [5] optimal designs for the two-parameter logistic function are constructed using four such criteria. Although the topic being addressed in this report is not the construction of optimal designs, it will be seen that the optimal augmentation of an existing design is closely related to optimal design methodology.

A. OPTIMALITY CRITERION

Perhaps the most widely used criterion for the construction of optimal designs is D-optimality. An n-point design is said to be D-optimal if the determinant of $(\underline{X}'\underline{H}\underline{X})^{-1}$, denoted $|(\underline{X}'\underline{H}\underline{X})^{-1}|$ is minimized for all n-point designs. Since $\underline{X}'\underline{H}\underline{X}$, often called the "information matrix," has determinant $|\underline{X}'\underline{H}\underline{X}| = 1/|(\underline{X}'\underline{H}\underline{X})^{-1}|$, the criterion can equivalently be expressed as maximizing $|\underline{X}'\underline{H}\underline{X}|$.

Dykstra [3] utilized the notion of D-optimality in augmenting an existing n-point design by specifying the $(n + 1)^{\underline{st}}$ point as the one which yields the largest increase in the determinant of the information matrix. (Dykstra expresses the information matrix as $\underline{X}'\underline{X}$ because his work is based on the assumption of ordinary least squares estimation, where the weight

matrix equals the identity matrix. Since it is appropriate to use weighted least squares estimation with the logistic function, the augmentation criterion used here is to maximize the increase in $|\underline{X}'\underline{HX}|$.) By subsequently considering the newly added point as part of the existing design, the process can be repeated to specify additional points.

B. IMPLEMENTATION PROBLEMS

One problem with the implementation of the augmentation criterion is that the level of the predictor variables must be controllable to some acceptable degree of accuracy in order to be able to make observations at the specified optimal new design points. A different sort of problem that can occur when $k > 1$ (that is, when there is more than one predictor variable) is that the optimal new design point may diverge to infinite values. In this case, the experimental region must be arbitrarily constrained to insure that the augmented design points have reasonable values. An example of this is given in Section V.B.

The particular form of the weight matrix gives rise to a third problem: the fact that $\underline{H}$ is a function of the parameters, which are not known at the time of experimentation, results in a "Catch-22" wherein it is necessary to know the parameter values in order to decide how best to estimate them! A practical solution to this dilemma is to use the existing experimental data to estimate $\underline{\beta}$ and then to use the estimate, $\underline{b}$, to specify an approximately optimal new design point. Such a procedure will be referred as an "approximate procedure" as opposed to the "exact procedure" which is only hypothetically possible.

Of course, it is best to update the parameter estimates before each

additional point is specified. However, it is often difficult, if not impossible, to do this because of time and cost constraints or because of the nature of the experiment. In this case, more than one additional point can be specified using the same estimate of $\underline{\beta}$ to estimate the information matrix at each stage. That is, the information matrix can be reestimated after each new design point is specified without reestimating the parameters. After, say, five new design points are specified in this way, the five new runs can be made and the new data used to update the estimates of $\underline{\beta}$ for specifying the next "block" of five points.

Note that within each block, the five new points are added one at a time; there may be a different set of five new points which would yield a larger increase in $|\underline{X}'\underline{\underline{H}}\underline{X}|$. Unfortunately, the calculations involved in maximizing the increase in $|\underline{X}'\underline{\underline{H}}\underline{X}|$ with respect to more than one new point quickly become intractable and the cost involved in a computerized numerical search approach is prohibitive.

C. EFFICIENCY

A measure of the amount of information in a design which does not depend on the number of points in the design, n, is the normalized determinant, $[1/n^{k+1}]|\underline{X}'\underline{\underline{H}}\underline{X}|$. (See Draper and St. John [2].) For example, when k = 1 D-optimal designs have $[1/n^{k+1}]|\underline{X}'\underline{\underline{H}}\underline{X}| = [P'Q' \ln(P'/Q')/\beta_1]^2$, where $P' = .824$, or $[1/n^2]|\underline{X}'\underline{\underline{H}}\underline{X}| = .0501185/\beta_1^2$. (See Kalish and Smith [5].) For the remainder of this section, only the one variable case will be discussed. By dividing the value of the normalized determinant for any particular design by $.0501185/\beta_1^2$, the efficiency relative to D-optimality is calculated. This will be called D-efficiency. (See Draper and St. John [1].)

Consider the problem of adding an n^{th} point to an existing (n - 1)-point design. If the parameter values were known, the n^{th} point could be specified using the exact augmentation procedure and the D-efficiency of the new design could be calculated. Denote the resulting efficiency DE_n. In practice only an approximately optimal n^{th} point can be specified. Letting DA_n denote the D-efficiency after adding the approximately optimal n^{th} point, a plot of DA_n versus n would illustrate the progress of an experimental design in achieving D-optimality. The ratio $R_n = DA_n/DE_n$, to be called relative efficiency, plotted against n would illustrate the progress of the approximate procedure in approaching the exact procedure.

Unfortunately, it is impossible to calculate any of these efficiencies in a practical situation since the true parameter values are not known. Therefore a simulation was performed to give a typical picture of what might be expected in an actual experiment.

IV. SIMULATION

A simulation study was conducted to evaluate the augmentation procedure under a variety of conditions in the one variable case. Six separate simulations (A1, A2, B1, B2, C1, C2) were run. In simulations A1 and A2, the initial design contained 25 points scattered between LD_{03} and LD_{73} while in simulations C1 and C2, the initial design contained 25 points, 20 of which were below LD_{05} or above LD_{95}. Simulations B1 and B2 represented a compromise between the A's and the C's.

The values of x_i for all three initial designs were the same, but their LD levels were controlled by changing the assumed parameter values. In Figure 1, the values of x_i are shown along with their corresponding probabilities, P_i, for each set of parameters. (The P_i's are calculated from model equation (1) or (4).) The three assumed parameter vectors are also shown in the figure.

A. PROCEDURE

In simulations A1, B1 and C1, new points were added in blocks of one (i.e., the parameters were reestimated after each new observation was specified and simulated), while in simulations A2, B2 and C2, the points were added in blocks of five.

Following is an outline of the simulation procedure:

(i) For each point in the initial design, an observation was simulated by generating a uniform random number, r_i, in the interval (0, 1). Each observation was defined as resulting in an "occurrence" or "nonoccurrence" by defining p_i such that

i	x_i	Simulations A1 and A2 $\underline{\beta} = \begin{pmatrix} -9.0 \\ 0.3 \end{pmatrix}$ P_i	Simulations B1 and B2 $\underline{\beta} = \begin{pmatrix} -30.0 \\ 1.1 \end{pmatrix}$ P_i	Simulations C1 and C2 $\underline{\beta} = \begin{pmatrix} -54.0 \\ 2.0 \end{pmatrix}$ P_i
1	18.8	0.034	0.000	0.000
2	22.0	0.083	0.003	0.000
3	22.6	0.098	0.006	0.000
4	23.5	0.125	0.015	0.001
5	23.7	0.131	0.019	0.001
6	24.3	0.153	0.037	0.004
7	25.2	0.192	0.093	0.027
8	25.5	0.206	0.125	0.047
9	26.9	0.283	0.399	0.450
10	27.2	0.302	0.480	0.599
11	27.6	0.327	0.589	0.769
12	28.1	0.361	0.713	0.900
13	28.3	0.375	0.756	0.931
14	28.6	0.397	0.811	0.961
15	28.8	0.411	0.843	0.973
16	29.7	0.478	0.935	0.996
17	30.9	0.567	0.982	1.000
18	31.2	0.589	0.987	1.000
19	31.3	0.596	0.988	1.000
20	31.7	0.625	0.992	1.000
21	31.7	0.625	0.992	1.000
22	32.4	0.673	0.996	1.000
23	32.6	0.686	0.997	1.000
24	33.0	0.711	0.998	1.000
25	33.5	0.741	0.999	1.000

Figure 1: Values of Parameter Vector ($\underline{\beta}$), Predictor Variable (x_i) and Response Probabilities (P_i) for Initial Designs Used in Simulation Study.

$$p_i = \begin{cases} 1 \text{ if } P_i \geq r_i \\ 0 \text{ if } P_i < r_i \end{cases}$$

where $P_i = \{1 + \exp[-(\beta_0 + \beta_1 x_i)]\}^{-1}$

(ii) The parameters were then estimated by inputting the simulated sample into a program developed by Jones [4] which solves for the maximum likelihood estimates. (An alternate program by Walker and Duncan [11], which uses the weighted least squares approach, would have yielded nearly the same estimates; the two methods are asymptotically equivalent.)

(iii) Using the estimated parameters, $\hat{\underline{H}}$ was calculated and a numerical technique was used to maximize $|\underline{X}'\hat{\underline{H}}\underline{X}|$ over choices of x_{n+1}. Here $\underline{X}$ is the $(n + 1) \times 2$ augmented design matrix with the $(n + 1)^{\underline{st}}$ row, $\underline{x}'_{n+1}$, representing the conditions for the approximately optimal new design point; $\hat{\underline{H}}$ is the $(n + 1) \times (n + 1)$ estimated weight matrix with $(n + 1)^{\underline{st}}$ weight, $\hat{P}_{n+1}\hat{Q}_{n+1}$. DA_{n+1} was calculated using the approximately optimal new design point as the $(n + 1)^{\underline{st}}$ point.

(iv) Step (iii) was repeated using the true parameter and weights. DE_{n+1} was calculated using the specified exactly optimal new design point as the $(n + 1)^{\underline{st}}$ point.

(v) The approximately optimal new design point (from step (iii)) was added to the design and its observation was simulated using the technique described in step (i). Note that the design was augmented

with the approximately optimal point but its observation was simulated using the true parameters. This corresponds to a real-life situation where the experimenter can only approximate the optimal new design point but the probability of response for that point follows the true (unknown) logistic function.

(vi) Simulations A1, B1 and C1 only: Treating the augmented design as the new "initial" design (resetting the sample size, n, to n + 1 each time a point was added), steps (ii)-(v) were repeated 24 more times so that the final design had 50 points.

(vii) Simulations A2, B2 and C2 only: Using the estimated parameters from (ii), steps (iii) and (iv) were repeated four more times by treating the augmented design as the new initial design (resetting the sample size, n, to n + 1 each time a point was added.) Thus, a new block of five points was added without reestimating the parameters.

(viii) Simulations A2, B2 and C2 only: Steps (ii)-(v) and (vii) were repeated four more times so that the final design had 50 points.

(ix) In order to estimate the expected values of DA_n and R_n, and to construct confidence intervals, steps (i)-(viii) were repeated 120 times.

B. RESULTS

For each simulation, expected values of DA_n and R_n were estimated by

averaging results from the 120 replications. Plots of DA_n versus n are shown in Figure 2 for simulations A1 and A2, in Figure 3 for simulations B1 and B2 and in Figure 4 for simulations C1 and C2. Also included are approximate 95% confidence intervals for n = 35 and 50. The analogous plots and confidence intervals for R_n are presented in Figure 5 for simulations A1 and A2, in Figure 6 for simulations B1 and B2 and in Figure 7 for simulations C1 and C2.

It seems reasonable to believe that the augmentation procedure would yield higher D-efficiencies when the parameters are reestimated before each new point is added than when points are added in blocks of more than one. However, in the cases considered (see Figures 2, 3 and 4) the size of the blocks did not have a significant effect on efficiency DA_n. This indicates that from a practical standpoint, very little is lost by augmenting in blocks of five rather than one. In addition, the high relative efficiencies in Figures 5, 6 and 7 reveal that the loss in efficiency due to the need to use estimated parameter values is minimal. Another observation to be made from Figures 5, 6 and 7 is that for poorer initial designs there seems to be a "start-up" phase before expected relative efficiency starts increasing.

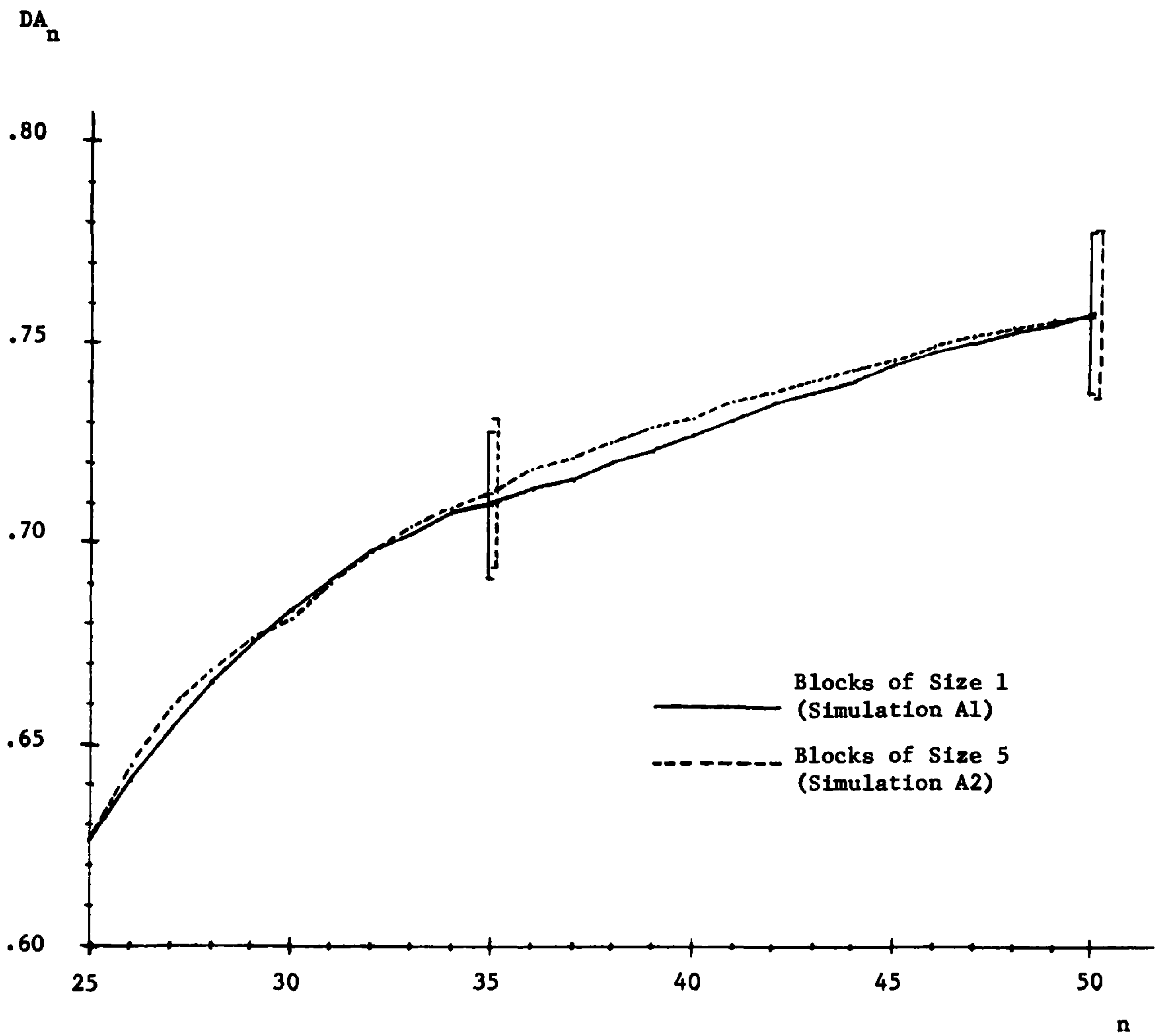


Figure 2: D-efficiency Versus Sample Size For Simulations A1 and A2; (β_0 = -9.0, β_1 = .30). 95% Confidence Intervals Shown for n = 35 and 50.

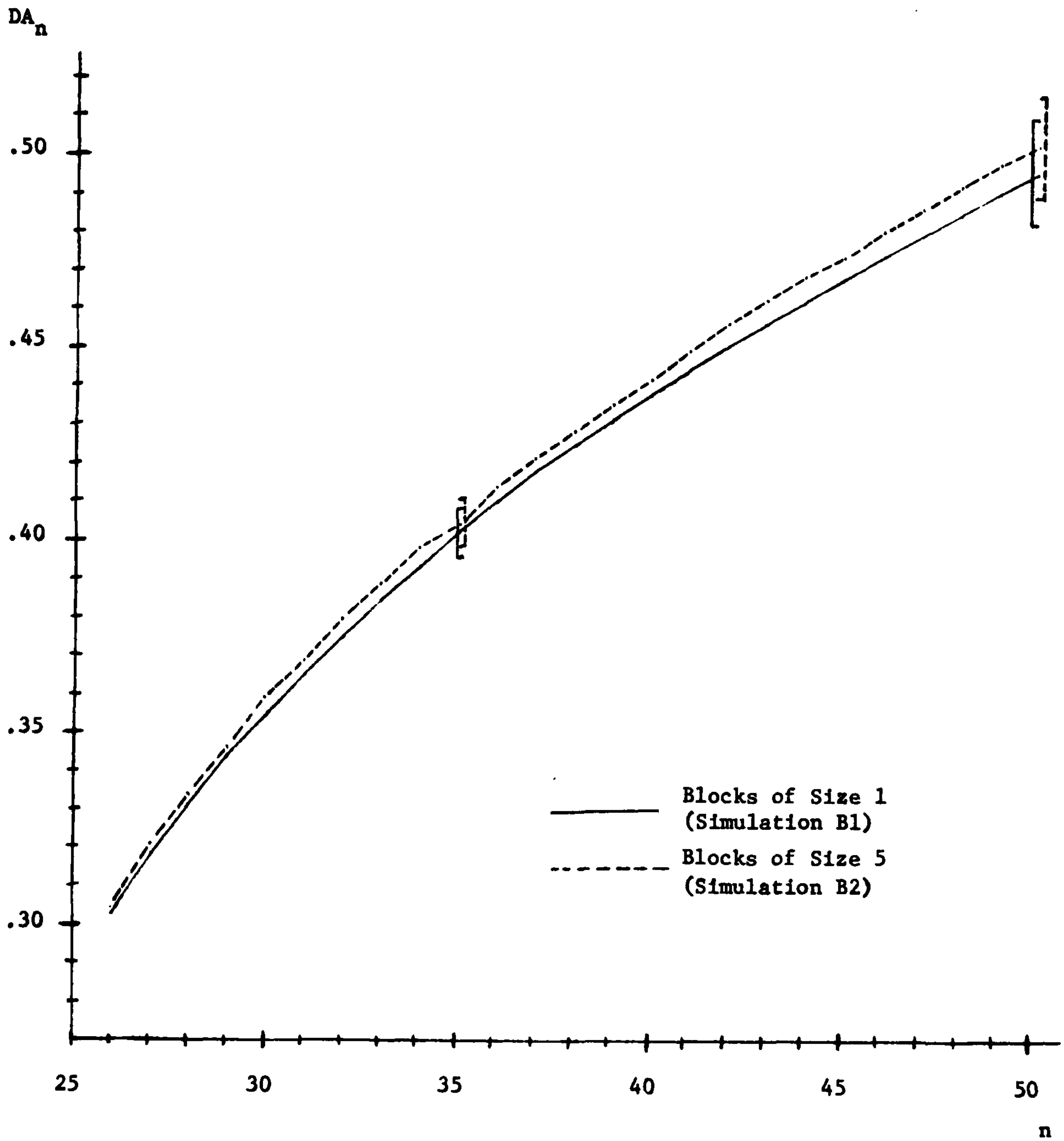


Figure 3: D-efficiency Versus Sample Size For Simulations B1 and B2; (β_0 = -30.0, β_1 = 1.1). 95% Confidence Intervals Shown for n = 35 and 50.

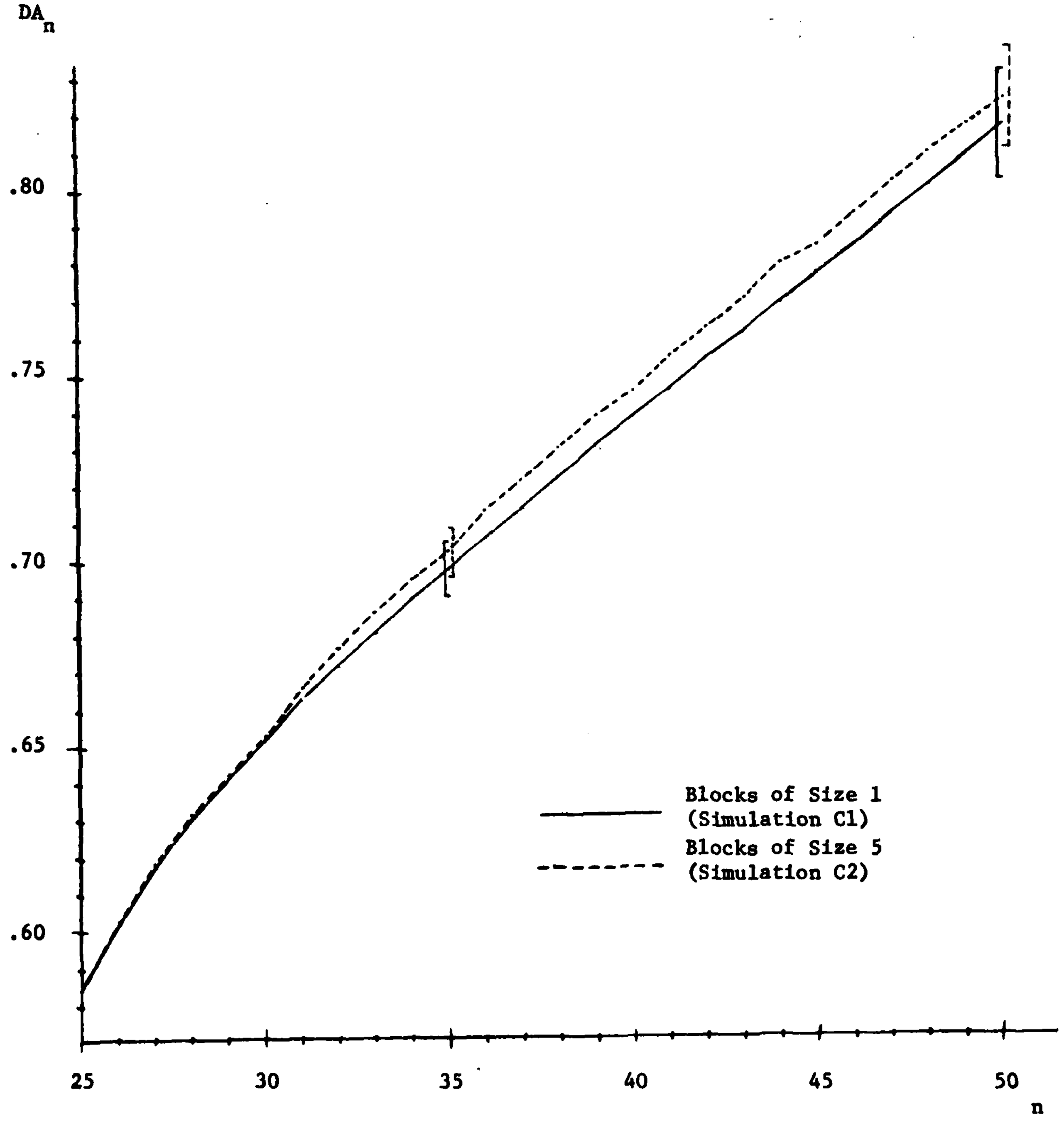


Figure 4: D-efficiency Versus Sample Size for Simulations C1 and C2; (β_0 = -54.0, β_1 = 2.0). 95% Confidence Intervals Shown for n = 35 and 50.

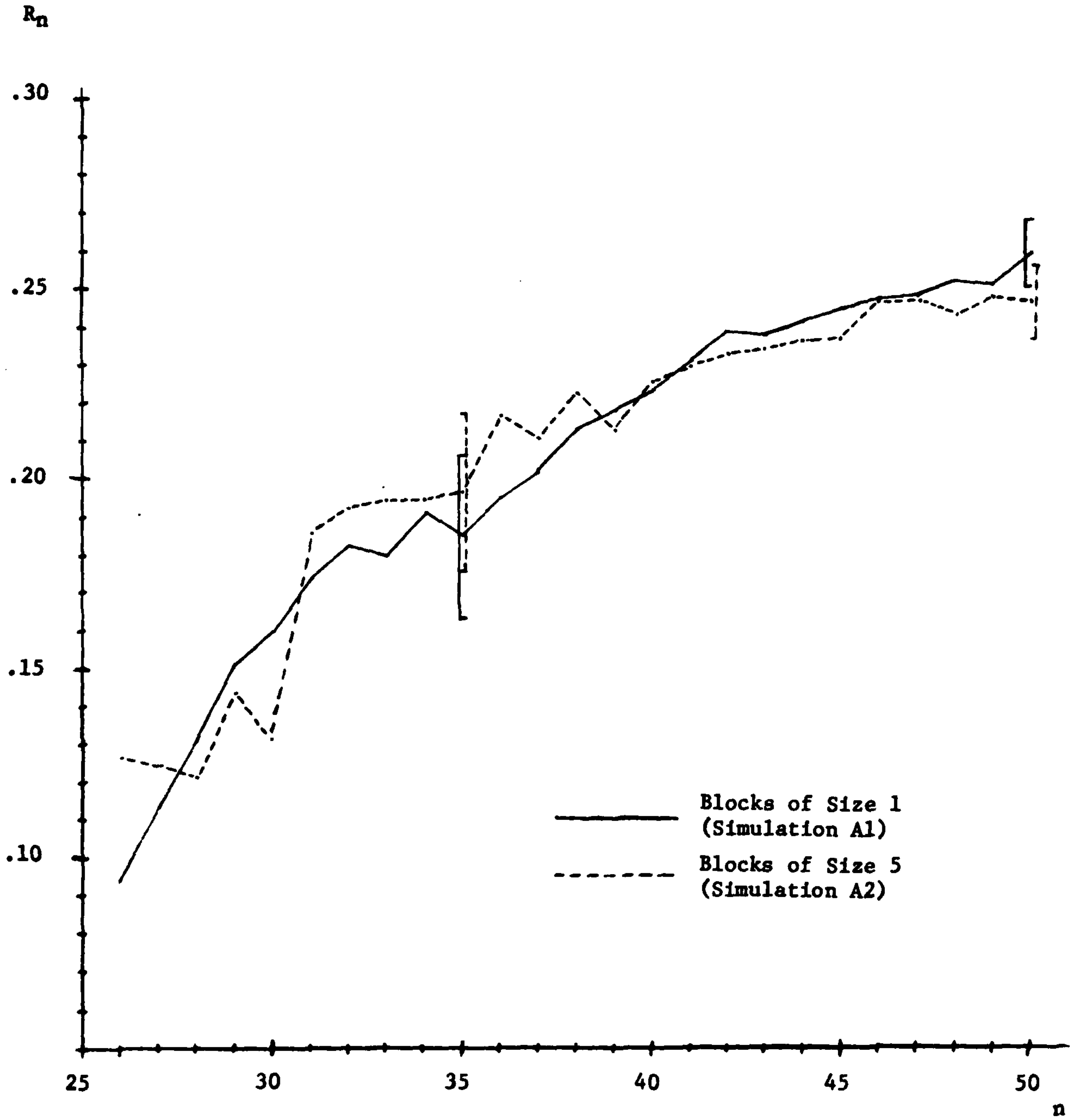


Figure 5: Relative Efficiency Versus Sample Size For Simulations A1 and A2; ($\beta_0 = -9.0$, $\beta_1 = .30$). 95% Confidence Intervals Shown for n = 35 and 50.

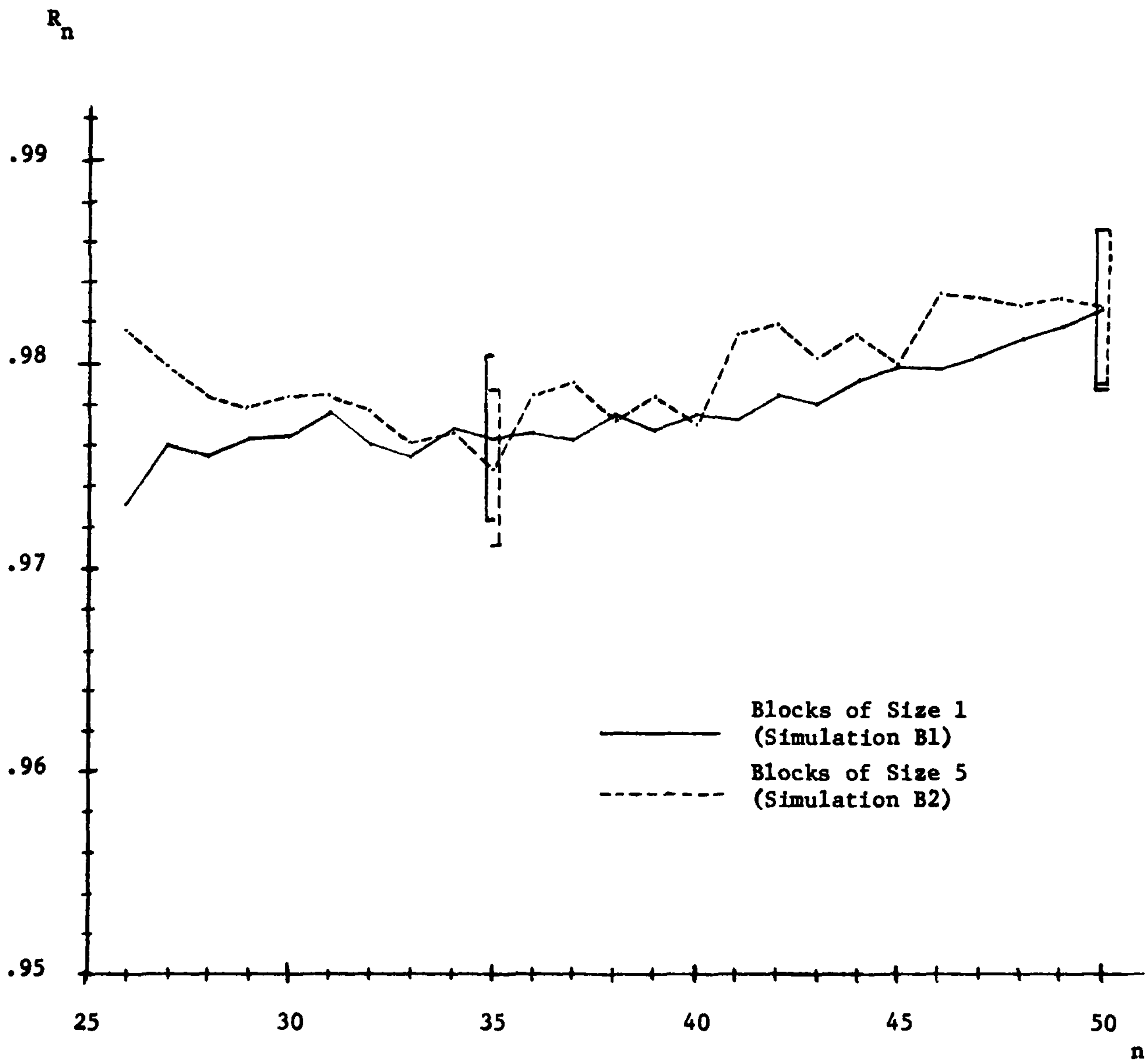


Figure 6: Relative Efficiency Versus Sample Size For Simulations B1 and B2; (β_0 = -30.0, β_1 = 1.1). 95% Confidence Intervals Shown for n = 35 and 50.

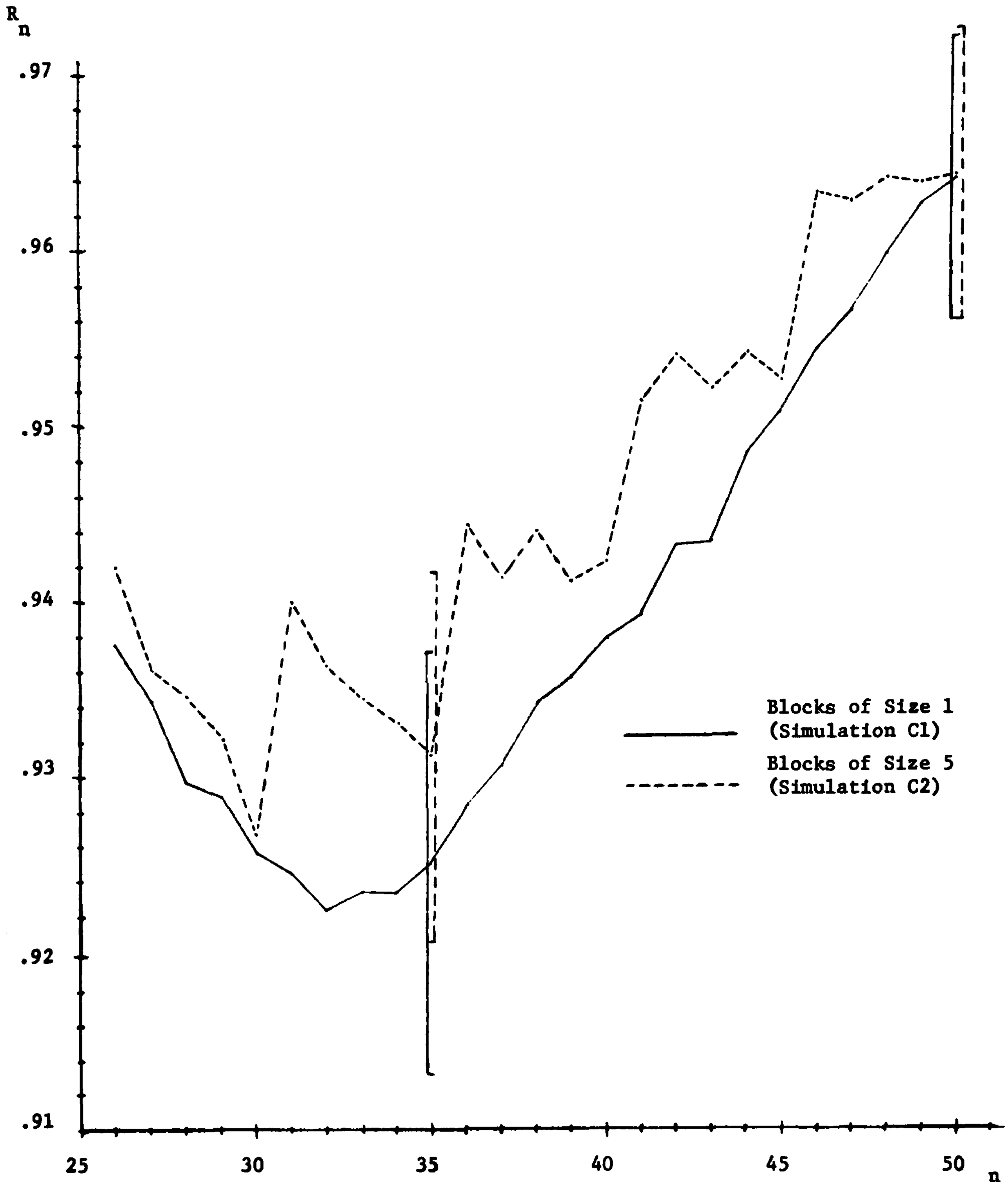


Figure 7: Relative Efficiency Versus Sample Size for Simulations C1 and C2; ($\beta_0 = -54.0$, $\beta_1 = 2.0$). 95% Confidence Intervals Shown for n = 35 and 50.

V. EXAMPLES

Three examples will be taken from the previously mentioned data set of twenty-eight $-G_x$ accelerator runs. (See [8] and [10].) In each case, the next five design points (one block of five) will be specified and the D-efficiencies after each new run will be estimated. If the experimenter chooses to work in blocks of size smaller than five (say three, for example), he or she would make only the first three runs specified and then stop to reestimate the parameters and specify three new runs.

A. TWO-PARAMETER CASE

Model A from [10] used peak force along the anatomical Z axis (denoted FHC) as the predictor of injury. The model, as estimated from the initial 28 runs, was given by

$$\hat{P}(\mathrm{FHC}) = (1 + \exp\{-[-7.3705 - .1072(\mathrm{FHC})]\})^{-1}$$

The existing design has an estimated D-efficiency of .2347 and the next five design points (one block of five) are listed in Figure 8(a) along with the estimated D-efficiencies after each point is added to the design.

For the second example, consider model 3 from [8] in which peak sled acceleration (denoted z_1) was the predictor of injury. The estimated model was given as

$$\hat{P}(z_1) = \{1 + \exp[-(-49.81 + .4472\, z_1)]\}^{-1}$$

For this case the existing design has an estimated D-efficiency of .0134 and the next five design points are listed in Figure 8(b) along with the estimated

$\underline{n}$	$\underline{FHC = x_n}$	Estimated D-efficiency
29	-78.9	.2592
30	-78.9	.2801
31	-78.9	.2977
32	-57.8	.3140
33	-80.4	.3298

Figure 8(a): One Block of Five Additional Design Points for Estimating Probability of Injury From Peak Force Along the Anatomical Z-Axis (FHC).

$\underline{n}$	$\underline{z_1 = x_n}$	Estimated D-efficiency
29	114.9	.0339
30	114.9	.0517
31	114.9	.0672
32	114.9	.0807
33	107.9	.0927

Figure 8(b): One Block of Five Additional Design Points for Estimating Probability of Injury From Peak Sled Acceleration (z_1).

D-efficiencies after each point is added to the design.

B. THREE-PARAMETER CASE

The following example will illustrate one approach for handling problems which arise in the three-parameter case. Model (7) from [8] used peak sled acceleration (z_1) and peak head angular velocity (x_3) to predict injury. The estimated model was given as

$$\hat{P}(z_1, x_3) = \{1 + \exp[-(-248.6 + 2.009\ z_1 + 0.11938\ x_3)]\}^{-1} .$$

As previously mentioned, when more than one predictor variable is being used, there is a need to restrict the design space so that the optimal points do not diverge to infinite values.

For this example, an assumption about the nature of the predictor variables was made which served to place probabilistic constraints on the design space. It was assumed that only the peak sled acceleration (z_1) is controllable and that peak head angular velocity (x_3) is related to z_1 by an equation of the form $x_3 = \alpha_0 + \alpha_1 z_1 + \varepsilon$ where ε has a normal distribution with mean 0 and variance σ^2. The estimates $\hat{\alpha}_0 = -2.57$, $\hat{\alpha}_1 = 1.312$, $\hat{\sigma}^2 = 3724.4$ were obtained from the data by least squares estimation.

To describe the augmentation procedure for this situation, it is convenient to introduce some additional notation. Define the function d(a, b) to equal the determinant of $\underline{X}'\underline{HX}$ after augmenting an existing design with new design point (a, b). Then the optimal value of z_1 is that value which maximizes $E[d(z_1, x_3)]$, where $x_3 = \alpha_0 + \alpha_1 z_1 + \varepsilon$. Here, the expectation is taken over ε; that is,

$$E[d(z_1, x_3)] = E[d(z_1, \alpha_0 + \alpha_1 z_1 + \varepsilon)] = \int_{-\infty}^{\infty} d(z_1, \alpha_0 + \alpha_1 z_1 + \varepsilon) f(\varepsilon) d\varepsilon,$$

where $f(\varepsilon) = [1/(2\pi)^{\frac{1}{2}}\sigma] \exp[-(\varepsilon^2/2\sigma^2)]$ is the probability density function of ε. After each new value of z_1 is specified, the expected new point $(z_1, E(x_3))$ is added to the existing design. Then the next point is specified using the newly augmented design as the "existing design."

After each new block of points is specified in this way, the new runs can be made. The resulting data is then used to replace the values of $E(x_3)$ with their actualizations and to update the estimates $\hat{\underline{\beta}}$, $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\sigma}^2$ and $\hat{\underline{H}}$.

Preliminary calculations for this particular example indicated that with $\hat{\underline{\beta}} = (-248.6, 2.009, 0.11938)'$, the function $\hat{P}(z_1, x_3)$ takes on values very close to 0.0 or 1.0 at all but an extremely narrow band of points in the z_1x_3-plane. As a result, a very high degree of precision needs to be carried through all calculations to avoid numerical inaccuracies. The narrow band of non-trivial points, the relatively large standard deviation about the estimated regression line $x_3 = \hat{\alpha}_0 + \hat{\alpha}_1 z_1$, and the high degree of precision required in specifying optimal new design points combine to make this example a poor illustration of the augmentation procedure. Thus, a different (fabricated) parameter vector $\hat{\underline{\beta}} = (-5.26, .02, .02)'$ was used for illustration purposes.

Using the fabricated parameter vector, the normalized determinant for the initial design is 1928.75. (Note that D-efficiency cannot be estimated because the D-optimal design for the three-parameter case is not known.) A block of five values of z_1 is listed in Figure 9 along with their corresponding values of $E(x_3) = -2.57 + 1.312\ z_1$ and the estimates of the new expected normalized determinants, $E[d(z_1, x_3)]$.

n	z_1	$E(x_3)$	Estimated Normalized Determinant
29	157.66	204.29	2201.66
30	157.07	203.51	2372.31
31	156.50	202.76	2499.36
32	155.96	202.05	2591.25
33	155.44	201.37	2654.74

Figure 9: One Block of Five Additional Design Points for Estimating Probability of Injury From Peak Sled Acceleration (z_1) and Peak Head Angular Velocity (x_3).

VI. SUMMARY

A procedure for optimal augmentation of an experimental design was presented. In particular, the augmentation procedure was applied to the problem of estimating the two-parameter logistic function. Since, for this case, the optimality criterion is dependent on the unknown parameter values, parameter estimates need to be used and updated periodically.

A simulation study was conducted to evaluate the augmentation procedure. It was found that updating the parameter estimates after each new design point is added does not result in significantly higher efficiencies than updating the estimates only after each block of five new points is added. It was also found that the loss in efficiency due to the need to use estimated parameter values rather than the true values is minimal.

VII. REFERENCES

[1] Draper, N. R., and St. John, R. C., "Models and Designs for Experiments With Mixtures: I, Background Material," Technical Report No. 360, Department of Statistics, University of Wisconsin, 1975.

[2] Draper, N. R., and St. John, R. C., "Designs in Three and Four Components For Mixture Models With Inverse Terms," Technometrics, Vol. 19, pp. 117-130, 1977.

[3] Dykstra, O. Jr., "The Augmentation of Experimental Data to Maximize $|\underline{X}'\underline{X}|$," Technometrics, Vol. 13, pp. 682-688, 1971.

[4] Jones, R. H., "Probability Estimation Using a Multinomial Logistic Function," J. Statist. Comput. Simul., Vol. 3, pp. 315-329, 1975.

[5] Kalish, L. A., and Smith, D. E., "Optimal Designs for Estimation of the Two-Parameter Logistic Function," Technical Report No. 112-4, Desmatics, Inc., 1980.

[6] Kiefer, J., "Optimum Experimental Designs," J. Royal Statist. Soc. - Ser. B, Vol. 21, pp. 272-319, 1959.

[7] Smith, D. E., "Research on Construction of a Statistical Model for Predicting Impact Acceleration Injury," Technical Report No. 102-2, Desmatics, Inc., 1976.

[8] Smith, D. E., "An Examination of Statistical Impact Acceleration Injury Prediction Models Based on $-G_x$ Accelerator Data from Subhuman Primates," Technical Report No. 102-6, Desmatics, Inc., 1978.

[9] Smith, D. E., and Gardner, R. L., "A Study of Estimation Accuracy When Using a Logistic Model for Prediction of Impact Acceleration Injury," Technical Report No. 102-5, Desmatics, Inc., 1978.

[10] Smith, D. E., and Peterson, J. J., "An Examination of Statistical Impact Acceleration Injury Prediction Models Based on Torque and Force Variables," Technical Report No. 112-1, Desmatics, Inc., 1979.

[11] Walker, S. H., and Duncan, D. B., "Estimation of the Probability of an Event as a Function of Several Independent Variables," Biometrika, Vol.54, pp. 167-179, 1967.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

14) TR-112-5

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
112-5

2. GOVT ACCESSION NO.
AD-A085 004

3. RECIPIENT'S CATALOG NUMBER
9

4. TITLE (and Subtitle)
OPTIMAL AUGMENTATION OF EXPERIMENTAL DESIGNS FOR ESTIMATION OF THE LOGISTIC FUNCTION.

5. ~~TYPE OF REPORT &~~ PERIOD COVERED
Technical Report,

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Leslie A./Kalish [redacted] Dennis E./Smith

8. CONTRACT OR GRANT NUMBER(s)
15) N00014-79-C-0128, N00014-75-C-1054

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Desmatics, Inc.
P. O. Box 618
State College, PA 16801

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR 207-037
NR 042-334

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Arlington, VA 22217

12. ~~REPORT DATE~~
11) May 1980

13. NUMBER OF PAGES
~~26~~ 12) 30

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Logistic Function
Estimation
Optimal Augmentation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A criterion for optimal augmentation of an experimental design is applied to the problem of estimating the logistic function. A simulation study is conducted to evaluate the procedure in the two-parameter case. Examples in the development of impact acceleration injury prediction models are given.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)